Optimal Combining Data for Improving Ocean Modeling

L.I.Piterbarg
University of Southern California
Center of Applied Mathematical Sciences
Kaprielian Hall, R108, 3620 Vermont Avenue
Los Angeles, CA 90089-2532
Phone (213) 740 2459, fax (213) 740 2424, e-mail piter@usc.edu
Award # N00014-09-1-0322
http : //www.onr.navy.mil/sci_tech/ocean/onrpgahm.htm

LONG-TERM GOALS

The long range scientific goals of the proposed research comprise: (1) developing rigorous approaches to optimal combining different kinds of observations (images, ADCP, HFR, glider, drifters etc) with output of regional circulation models for accurate estimating the upper ocean velocity field and mixing characteristics (2) constructing computationally efficient and robust estimation algorithms based on alternative parameterizations of uncertainty and comprehensive testing them on synthetic data (3) processing real data in the Adriatic and Ligurian Sea (MREA coastal experiments) via new techniques

OBJECTIVES

The objectives for the second year of research were:
- Enhancing the fusion method for computing surface velocities [1] by accounting for uncertainties in tracer generation and dissipation.
- Constructing and testing fusion algorithms for data coming from different sources at different resolution.
- Developing fusion methods based on the fuzzy logic [2-4] for estimating oceanic parameters from small biased samples.

APPROACH

We develop theoretical approaches to the data fusion problem in context of the possibility theory (fuzzy logic) and in the framework of the classical theory of random processes and fields covered by stochastic partial differential equations. We also design computational algorithms derived from the theoretical findings. A significant part of the algorithm validation is their testing via Monte Carlo simulations. Such an approach provides us with an accurate error analysis. Together with my collaborators from Rosenstiel School of Marine and Atmospheric Research (RSMAS), Consiglio Nazionale delle Ricerche (ISMAR, LaSpezia, Italy), Naval Research Laboratory (Stennis Space Center, Mississippi), ENEA (Rome, Italy), Koc University (Istanbul, Turkey), and Naval Postgraduate School (Monterrey, CA) we implement the algorithms in concrete ocean models such as QG, POM, MICOM , NCOM, and MFS, as well as carry out statistical analysis of real data sets by means of new methods.

Report Documentation Page					*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE **2010**		2. REPORT TYPE			3. DATES COVERED **00-00-2010 to 00-00-2010**
4. TITLE AND SUBTITLE **Optimal Combining Data for Improving Ocean Modeling**					5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)					5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Southern California,Center of Applied Mathematical Sciences,Kaprielian Hall, R108, 3620 Vermont Avenue,Los Angeles,CA,90089-2532**					8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)					10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**7**	

WORK COMPLETED

1. Enhancing the fusion method developed by PI for computing surface velocities by accounting for uncertainties in tracer generation and dissipation.

A newer version of the method for optimal estimating surface velocities from model output and tracer observations [1] has been developed on the base of representing uncertainties via possibility theory means [2-4]. The focus was on accounting for uncertainties in generation and dissipation. In particular, the approach targeted color images formed mostly by chlorophyll concentration for which production and dissipation processes are difficult to parameterize and quantify. The dependence of the estimation error on parameters of the possibility distributions involved was comprehensively studied [5]. Explicit estimation formulas were obtained not involving any numerical integration and optimization which is of great importance for fast computations when the number of grid points is very large.

Extended Monte Carlo simulations were carried out to evaluate the estimation skill on different model flows in conditions of poorly known the right hand side (RHS) of the tracer transport equation. Different scenarii for uncertainties in RHS were examined and recommendation for choosing parameters of the corresponding possibility distributions (membership functions) were worked out.

2. Constructing and testing fusion algorithms for data coming from different sources at different resolution.

A general approach to fusing data at different resolution has been developed by combining ideas from two novel areas: multi objective optimization and fuzzy set theory. The focus was on estimating (Eulerian) velocities from coarse direct observations and fine tracer data.

Typically, direct measurements covering a significant area cannot provide a high degree of the space resolution because of technical difficulties and high cost of such measurements. For example the distance between ADCP (Acoustic Doppler Current Profiler) stations in NAVOCEANO shipboard surveys (California Coastal Region) is about 18.5 km. On the other hand the space resolution of satellite observations of the sea surface temperature and chlorophyll concentration is much higher, 1-2 km or less depending on a type of radiometers.

We have constructed and tested an algorithm for optimal combining coarse irregular ADCP measurements with tracer observations, either SST or color, on a fine grid. Besides multi objective optimization ideas the algorithm takes advantage of parameterizing uncertainties in models and observations via the fuzzy logic approach which is more general than the classical statistics paradigm. The method was comprehensively tested on synthetic velocity fields and compared to traditional interpolation procedures.

3. Developing fusion methods based for estimating oceanic parameters from small biased samples.

Sparse observations, biasness, and small samples pose serious obstacles for application of classical statistical methods in processing ocean data. An alternative approach has been developed for fusing such data and making inferences base on the fundamental concept of Pareto optimal solutions in multi objective optimization problems [6].

First, a fuzzy estimator based on a triangle membership function was suggested and studied for an ocean parameter, such as temperature or velocity at a fixed point, from two small biased samples [7]. Its performance was compared with the weighted mean and weighted median for different types of the noise distribution.

Then, numerical simulations were carried out to compare the suggested estimator with those based on other types of membership functions.

Finally, the structure of the Pareto optimal set for the number of sensors (information sources) more than two has been studied and the corresponding aggregated estimators were examined by Monte Carlo means.

RESULTS

1. The developed method proved to be accurate and extremely computationally efficient in conditions of poor known RHS for the tracer. Improvement comparing to the model output varied from 8 to 27 % depending on the considered flow, parameters of the tracer patch and uncertainties in RHS. One of the examples is presented in Fig.1 where the 'true' velocity field was represented by a circulation cell superimposed by eddies and the model output was represented by the circulation cell itself.

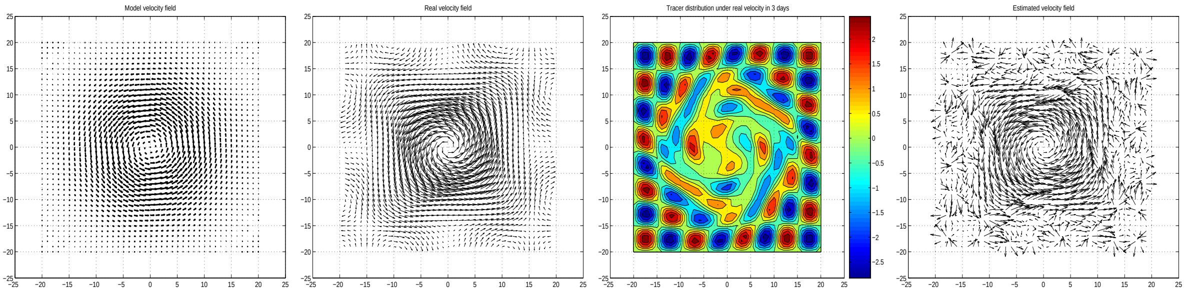

Figure 1. Fusion of tracer observations and circulation model output (experiment with a circulation cell superimposed by eddies): 1) Background (model output). 2) 'Truth'. 3) Tracer observations (color). 4) Estimate.

The most important finding is that the method performs better than its predecessors developed in [1] in the case of significant uncertainties in tracer generation and dissipation as proven by Monte Carlo simulations and a theoretical error analysis. We also established a low sensitivity of the method performance to the location and initial size of the tracer patch. The most influential parameter is the forcing intensity in the transport equation, its high values do not allow to improve the estimate comparing to the model output.

2. The main result in data fusion at different resolution is that the developed method clearly outperforms traditional interpolation procedures which are intrinsically not able to capture subgrid scales. One of the experiments is illustrated in Fig. 2 for a gyre superposed with a periodic system of eddies. As one can see the method allows for very accurate estimate of the complex circulation patterns, while the classical interpolation approach provides an estimate of the background circulation only(large scale gyre).

Another example deals with a hypothetical coastal current perturbed by an isolate eddy (Fig.3). Clearly, coarse velocity measurements are not able to detect the eddy via interpolation. At the same time combining velocity data with tracer measurement on a fine grid one can get a clear idea on the size and location of the eddy.

In summary, the improvement achieved by using the suggested estimator comparatively to the optimal interpolation turned out to be in the range $10\% - 48\%$ depending mostly on the flow characteristics and the ratio of the space steps in velocity and tracer observations.

3. As for the estimation problem for small biased samples, the most principal conclusion is that the developed fuzzy estimator performs better than traditional statistical tools in a majority of scenarii with two biased small samples. Our approach models a typical situation in estimating the ocean state when information comes from different sources. The estimation skill was quantified by the dependence of the estimation error on two parameters α and β characterizing the ratio of biases and noise intensities for two sources. The dark area in Fig. 4 corresponds to the set of α and β where the fuzzy estimator is better than the traditional weighted mean or median widely used in

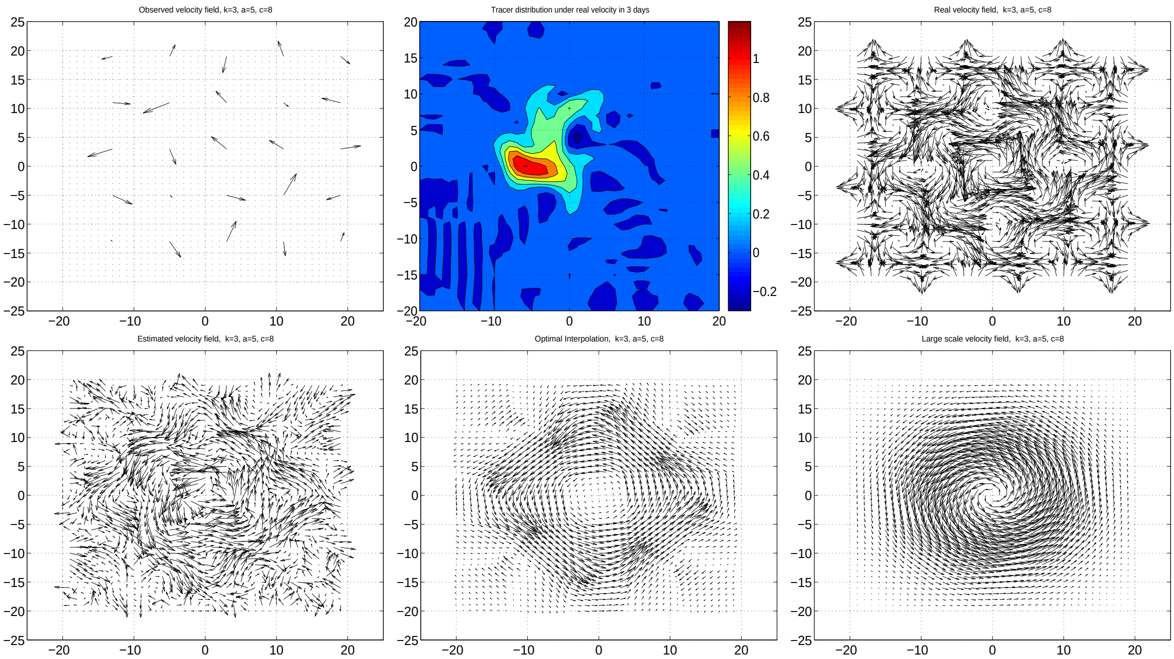

Figure 2. Fusion of synthetic ADCP and tracer data (a gyre superposed with a periodic eddy system). Upper panel: 1) Velocity observations. 2) Tracer observations. 3) 'True' velocities. Bottom panel: 1) Velocities estimated by fuzzy multi-objective algorithm . 2) Velocities interpolated by a standard interpolation scheme. 3) Large scale component of 'true' velocity field (background) days.

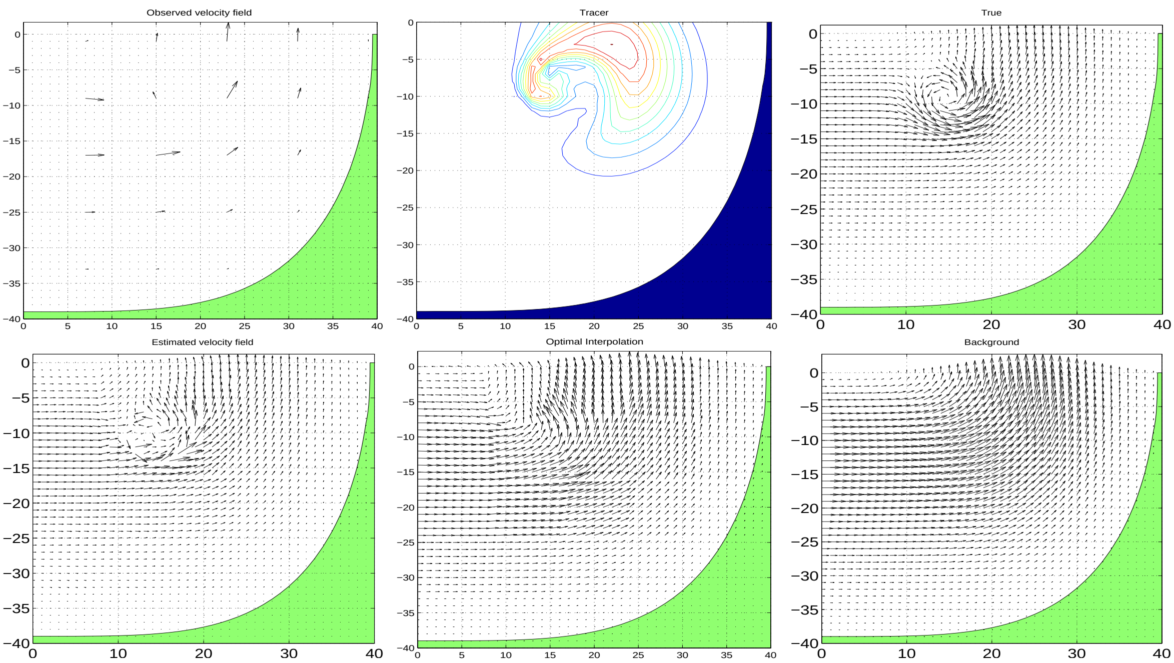

Figure 3. Fusion of synthetic ADCP and tracer data for a coastal flow with an isolated eddy. Upper panel 1) Velocity observations. 2) Tracer observations 3) 'True' velocities. Bottom panel 1) Estimated velocities . 2) Interpolated velocities. 3) Large scale component of 'true' velocity field (background).

oceanographic and atmospheric studies, [8]. As one can see the fuzzy estimator provides the less error in terms of both, bias and standard error, for the most of values of α and β

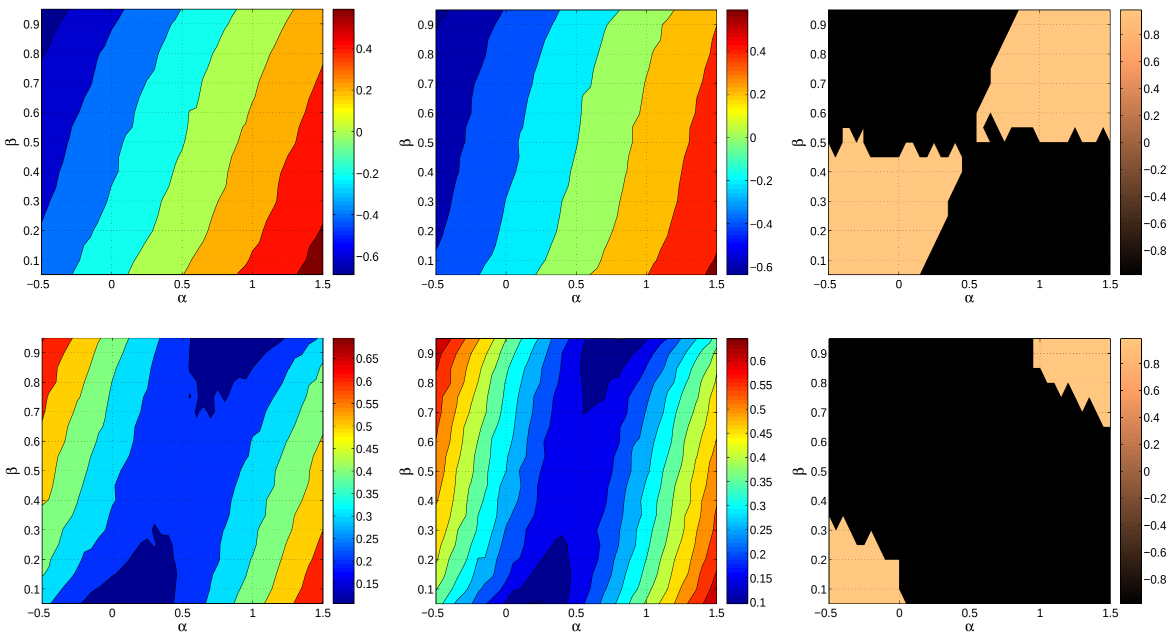

Figure 4. Comparison of the multi objective fuzzy estimate with weighted median (WMED) for Cauchy samples. Top panel: Bias. 1) WMED, 2) Fuzzy, 3) The dark area indicates the set of parameters α, β, where the fuzzy estimate is better than WMED in terms of bias. Bottom panel: Standard Error 1) WMED, 2) Fuzzy, 3) The dark area indicates the set of parameters α, β, where the fuzzy estimate is better than WMED in terms of STD.

In general it was proven by using Monte Carlo simulations and some theoretical means that the suggested approach improves estimates comparatively to traditional statistical methods like the weighted mean (optimal interpolation) and the weighted median, especially for distributions with heavy tails. The fuzzy estimator is also more efficient when noise distributions are different for different sources.
In summary, our theoretical findings and simulation results might serve as a guidance for choosing a particular estimation method from the discussed ones based on preliminary information on relations between unknown biases and variances in observations and model output.

IMPACT/APPLICATIONS

The developed data/model fusion methods are highly portable and computationally efficient, making them very valuable in the framework of operational strategies for rapid assessment and quick response. They have therefore some significant advantages with respect to other techniques requiring complete assimilation of the tracer information in the dynamical velocity models. These techniques, even though more powerful, require significant coding and computational time and they have to be set up in advance for the specific operational model in use. Moreover, the developed methods are resistant to uncertainties in tracer generation/dissipation , are capable to aggregate data at different resolutions, and account for sample biasness. Thus, we expect that our results will stimulate more efforts in developing fusion methods which carry no risk of ruining a model during the running time and , in addition, are well theoretically founded.

As for applications, both methods contribute to a variety of practical environmental and operational issues such as monitoring and forecasting pollutant spreading, search and rescue operations in the sea, and prediction of fish larvae.

TRANSITIONS

The developed velocity fusion algorithm was used in RSMAS and ENEA (La Spezia, Italy) to test it in the operational Mediterranean Forecasting System (MFS).

RELATED PROJECTS

1. "Predictability of Particle Trajectories in the Ocean", ONR, PI T.Ozgokmen , RSMAS, N00014-05-1-0095
2. "Lagrangian turbulence and transport in semi-enclosed basins and coastal regions", ONR, PI A Griffa, RSMAS, N00014-05-1-0094

REFERENCES

1. L.I. Piterbarg, (2009), A simple method for computing velocities from tracer observations and a model output, Applied Mathematical Modeling, 33, 36933704
2. D.Dubious and H. Prade, (1986), Possibility theory, Plenum Press, New York and London,
3. D. Dubois, H.Prade, and R.R. Yager, Fuzzy Information Engineering, 1997, John Wiley & Sons, Inc.
4. G. Shafer,(1976), A mathematical theory of evidence, Princeton University Press.
5. L.I. Piterbarg and L. Ivanov, (2010), Fuzzy-logic based algorithm for estimating circulation patterns, Fuzzy Sets ans Systems, submitted
6. Deb, K., (2001), Multi-Objective Optimization Using Evolutionary Algorithms, Willey.
7. L.I.Piterbarg, (2010), Parameter estimation from small biased samples: statistics vs fuzzy logic, Fuzzy Sets and Systems, under revision
8. Ghil, M., (1989), Meteorological Data Assimilation for Oceanographers. Part 1; description and theoretical framework, Dyn. Atm. Oceans, 13, 3-4, 171-218.

PUBLICATIONS

1. L.I. Piterbarg, (2009), A simple method for computing velocities from tracer observations and a model output, Applied Mathematical Modeling, 33, 3693-3704
2. L.I. Piterbarg, (2010), Computing Lagrangian statistics from tracer observations and a model output, Applied Mathematical Modeling, v.34, 11, 3674-3684
3. A. Mercatini , A. Griffa , L. Piterbarg, E. Zambianchi , M.G. Magaldi, (2010), Estimating surface velocities from satellite data and numerical models: Implementation and testing of a new simple method, Ocean Modeling, 33, 190-203
4. L.I.Piterbarg, (2010), Parameter estimation from small biased samples: statistics vs fuzzy logic, Fuzzy Sets and Systems, under revision
5. L.I. Piterbarg and L. Ivanov, (2010), Fuzzy-logic based algorithm for estimating circulation patterns, Fuzzy Sets ans Systems, submitted
6. M. Caglar, T.Bilal, L.I. Piterbarg, (2010), Lagrangian prediction and correlation analysis with Eulerian data, *Turkish J. Earth Sciences*, accepted

7. L.I. Piterbarg, (2009), Particle dispersion in stochastic flows with linear shear, LAPCOD meeting, La Londe-les-Maures 2009, Meeting Abstracts
8. E. Zambianchi, A. Mercatini, A. Griffa, L. Piterbarg, M. Magaldi, (2009), Implementation of a new simple method to estimate surface velocities from satellite data and numerical models, LAPCOD meeting, La Londe-les-Maures 2009, Meeting Abstracts
9. L.I. Piterbarg, (2009), Estimating Lagrangian trajectories from tracer observations and model output, LAPCOD meeting, La Londe-les-Maures 2009, Meeting Abstracts